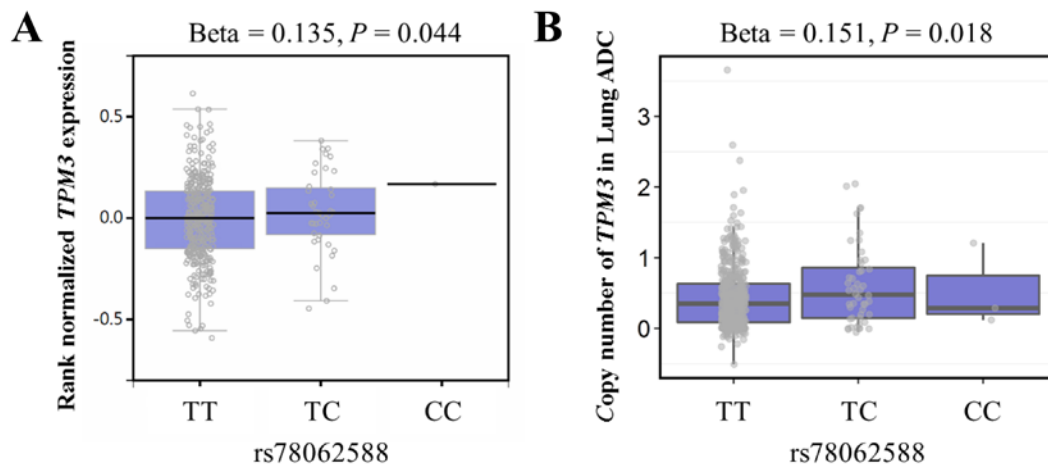
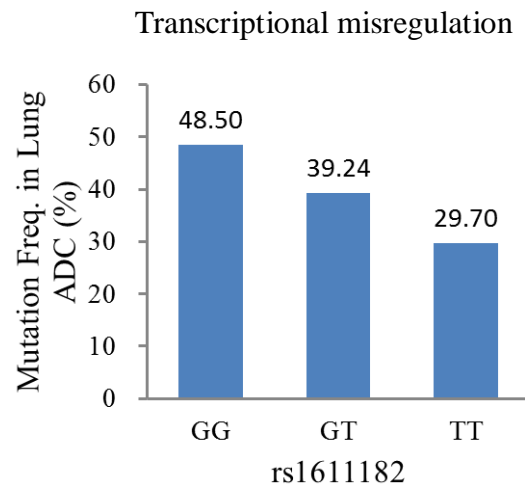


Supplementary Figure S1. A flow chart of statistical analysis in overall lung cancer. First, a total of 348 lung cancer driver genes (lung CDGs) were collected from the COSMIC Cancer Gene Census, TCGA LUAD&LUSC projects, and the IntOGen database (Supplementary Table S2). Then we identified 139,666 SNPs that were associated with driver gene expression (eSNPs) from GTEx v6p release as well as 2,041 nonsynonymous variants in lung CDGs using Variant Effect Predictor (VEP). In the discovery stage, we tested the associations between selected SNPs and overall lung cancer risk in the OncoArray dataset. SNPs with $p < 0.001$ were further assessed in the DCEG Lung Cancer Study. As a result, 234 SNPs that mapped to five lung cancer driver genes were identified. A series of bioinformatics analysis were conducted for significant variants.



Supplementary Figure S2. Associations of rs78062588 with *TPM3* expression and *TPM3* copy number. (A) Rank normalized *TPM3* expression in normal lung tissues with relation to genotypes of rs78062588. The association between rs78062588 and rank normalized expression of *TPM3* was investigated using linear regression model. The slope (Beta) and P value of the linear regression model were from the GTEx v6p release (1). (B) Association between rs78062588 and somatic copy number alterations of *TPM3* in lung adenocarcinomas (lung ADC). Data were from the Cancer Genome Atlas (TCGA) (2,3). Linear regression model was used to assess the association between rs78062588 and somatic copy of *TPM3*. Age, gender, smoking status, clinical stage, and the first ten principal components were adjusted as covariates. The slope (Beta) and P value were derived from the linear regression model. The box plots display the first and third quartiles (top and bottom of the boxes), the median (band inside the boxes), and the lowest and highest point within 1.5 times the interquartile range of the lower and higher quartile.



Supplementary Figure S3. Genotype of rs1611182 and frequency of truncation mutations within transcriptional misregulation pathway in lung ADCs. Freq.: frequency; Lung ADC: lung adenocarcinoma.

References

1. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **2013**;45:580-5.
2. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **2012**;489:519-25.
3. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **2014**;511:543-50.